Statistical Shape Modelling provides a responsive measure of morphological change in knee osteoarthritis over 12 months

Jennifer S Gregory[1†], Rebecca J Barr[1,2†], Kanako Yoshida[1], Salvatore Alesci[3], David M Reid[1], Richard M Aspden[1]

[†]These authors contributed equally to this paper

[1]School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, UK.

Current affiliations: [2]Medicines Monitoring Unit (MEMO), Division of Molecular & Clinical Medicine, School of Medicine, University of Dundee; [3]Takeda Pharmaceuticals, Washington District of Columbia, USA.

Corresponding author
Professor R.M. Aspden
Institute of Medical Sciences
University of Aberdeen
Foresterhill
Aberdeen
AB25 2ZD
Email: r.aspden@abdn.ac.uk

ABSTRACT

Objectives: Responsive biomarkers are needed to assess the progression of osteoarthritis (OA) and their lack has hampered previous clinical trials. Statistical Shape Modelling (SSM) from radiographic images identifies those at greatest risk of fast-progression or joint replacement, but its sensitivity to change has not previously been measured. This study evaluates the responsiveness of SSM in knee OA in a 12-month observational study.

Methods: 109 people were recruited, who had knee radiographs in the previous 12 months, and grouped based on severity of radiographic OA (Kellgren Lawrence grading). An SSM was built from three dual-energy x-ray absorptiometry scans at 6-month intervals. Change-over-time and OA were assessed using generalised estimating equations, Standardized Response Means (SRM) and Reliable Change (RC) Indices

Results: Mode 1 showed typical features of radiographic OA and had a strong link with KLG but did not change significantly during the study. Mode 3 showed asymmetrical changes consistent with medial cartilage loss, osteophytes and joint malalignment and was responsive to change, with a 12-month SRM of 0.63. The greatest change was observed in the moderate radiographic OA group (SRM 0.92) compared with the controls (SRM 0.21) and the RC index identified 14% of this group whose progression was clinically significant.

Conclusions: Shape changes linked the progression of osteophytosis with increasing malalignment within the joint. Modelling of the whole joint enabled quantification of change beyond the point where bone-to-bone contact has been made. The knee SSM is, therefore, a responsive biomarker for radiographic change in knees over 12 months.

Keywords: Knee osteoarthritis, Statistical shape modelling, Kellgren-Lawrence grading, Imaging biomarker, reliable change

Key Mesages

- Statistical shape models of DXA knee images identified OA progression over 12 months
- Greatest changes in shape scores were in individuals with moderate OA at baseline (KLG 2).
- Quantitative modelling of the whole knee provides a biomarker for radiographic change over 12 months.

Introduction

Although many measures of radiographic osteoarthritis (OA) have been developed, few are sensitive enough to identify changes within one year. Many relate to a single defined feature, *e.g.* joint space width (JSW) in radiographs [1-3] or cartilage thickness or volume from MRI [4-6], and do not reflect the full extent of structural changes in the joint. Currently no pharmaceutical trials of Disease-modifying OA drugs (DMOADs) have gathered sufficient evidence to be granted approval for use in OA. There is a pressing need for sensitive biomarkers that capture changes in multiple features of radiographic OA to adequately power future clinical trials. These biomarkers also have to be well-validated, qualified and accepted by regulatory bodies, which makes the process of identifying biomarkers very challenging.

A biomarker's responsiveness is often measured using the Standardized Response Mean (SRM), the difference between the mean values at two timepoints divided by the standard deviation of the change, with higher values indicating a more responsive measure [7]. Using the SRM enables a comparison to be made between measures on totally different scales. While the SRM measures overall change in a whole group, the reliable change index (RC) was devised to identify whether the observed change in an individual is statistically meaningful in the presence of measurement error [8]. There are a number of different approaches to calculating the RC and these are described in more detail by Hinton-Bayre [9, 10]. This approach has recently been explored for JSW [11] and we adopt the same method here. In brief, the RC index for an individual is defined by the difference between their first and second measurements divided by a standard error associated with the measurements [9].

In recent years, research into the use of Statistical Shape Modelling (SSM) to measure joint morphology has shown its potential as an imaging biomarker for osteoarthritis. There have been a number of studies in the hip that have captured both the characteristic deformation of the femoral head that occurs during OA [12-14] as well as the shapes that indicate a higher risk of incident radiographic OA or Total Hip Replacement (THR) [15-18]. Few studies however have examined knee shape in OA.

Knee SSMs can and have been built using a range of imaging modalities, including MRI and radiographs. Three-dimensional knee models from MRI images [19, 20] have provided valuable insight into localised changes resulting from OA. In contrast 2D models, whether radiographic or from pQCT slices [21, 22], are easier to visualize. Dual-Energy X-ray Absorptiometry (DXA) is an imaging modality that produces radiograph-like images for accurate measurement of bone mineral density (BMD) and the diagnosis of osteoporosis. This modality has been little explored for assessment of OA.

The aim of this study was to test whether SSM is sensitive enough to measure changes in knee morphology over 6-12 months in a mixed-sex, observational cohort with a range of severities of radiographic knee OA at baseline using sequential Dual Energy X-ray Absorptiometry (DXA) images.

SUBJECTS AND METHODS

Subject recruitment

The Aberdeen Hip And Knee Study (AHAKS) is a prospective study using subjects recruited from the Grampian NHS Radiology Information System (RIS). Computerised searches of the database identified subjects aged >30 years who had undergone bilateral radiographs of their knees in the previous 12 months. Radiographic reports were examined by a clinician to assess suitability for the study. Subjects were excluded based on the following criteria: surgical interventions (including joint prostheses and osteotomies), inflammatory arthropathies, congenital/developmental dysplasias, metabolic bone disease or absence of a formal report on the Radiology Information System.

Following subject identification, letters were sent to the referring physician to seek their help in recruiting the subject into the study (no incentive was offered). The referring physician sent an information pack to the subject who was asked to complete a form and return it to indicate interest in participating. The Grampian Research Ethics Committee approved the study (reference 06/S0801/116) and each subject's written, informed consent was obtained when they attended for scanning, in accordance with the declaration of Helsinki.

Radiographic grading

The radiographs of patients who participated in the study were scored for both knees by a single reader blinded to clinical diagnosis using the Kellgren-Lawrence grading (KLG) system [23, 24]. Subjects were classified into **control**: KLG 0 in at least one knee and no more than KLG 1 in the other, **doubtful OA** had KLG 1 in both knees, **moderate OA** had KLG 2 in one or both knees, and those classed as **severe OA** had at least one knee with KLG 3 or4 based on the KLG of their more severe joint. Subjects underwent Dual Energy X-ray absorptiometry (DXA) scans (iDXA, GE Lunar) of both knees at three visits (baseline, 6 and 12 months) using standard positioning by qualified DXA radiographers.

Repeatability of KL grading was calculated for 50 knees from matched baseline radiographic and DXA images from 25 people. Inter- and intra-observer Quadratic Weighted Kappa (QWK) values were calculated for (A) repeated grading of DXA scans and (B) radiograph vs. DXA

scans for 3 observers (KY, SG-S and DMR) which were presented independently and in random order with at least 1 week between repeated gradings of the same image. Grading was mostly classed as "Almost-perfect" using the Landis-scale [25] with one pair of observers falling into the "substantial agreement" range (<0.8) as QWK values ranged from 0.89-0.94 for intra-observer and 0.79-0.85 for inter-observer variability. Almost identical results were observed for agreement between the matched iDXA and radiograph images (QWK 0.83-0.87 intra-observer agreement).

Statistical Shape Modelling

Details of the statistical shape modelling technique (SSM) used have been described elsewhere [26] and more details are provided in Supplementary Data S1. Briefly, SSM uses a set of points to describe the outline of an object. Each point refers to the same location in every image. The coordinates of these points were subjected to Procrustes transformation to scale, rotate and translate them, so removing influences of overall size. Principal components analysis was then applied to the coordinates of the set of points for all the images. This is a 'data reduction' approach that produces a small set of orthogonal 'modes' (principal components) from the large number of coordinates. Each mode describes variations in shape that occur in a coordinated fashion. Raw mode scores for the set of images were normalised to have zero mean and scaled so that variation in each image is expressed in units of standard devations. Modes are ranked in descending order of variance explained and this was visualised using a scree plot (Supplementary Figure S1) [27] to help to decide how many modes should be included in further analysis.

Statistical Shape Modelling was performed using the active shape modelling toolkit from the University of Manchester for point placement; a software program that runs within MATLAB (The Math Works Inc, Natick, United states) software environment (http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/software/am_tools_doc/index.html); Custom-made software was used for model construction and analysis (SHAPE, Aberdeen University, Aberdeen, UK). For point placement, all images of the same knee taken during the study were displayed in a random order. In this study we built an SSM of the knee in which 84 points were placed around the outlines of the features of interest (Figure 1). This model included the distal femur, proximal tibia and osteophytes, allowing visualization of common radiographic features observed in OA. Images of both knees at each timepoint were included in the model.

Statistical Analysis

Results are shown as mean with associated (standard deviation) or [95%CI]. Comparison between the groups at baseline was performed using one-way ANOVAs for continuous variables (age and BMI), or Kruskal-Wallis ANOVA on Ranks if data were not normally distributed using the Shapiro-Wilk test, and a Chi-squared for differences between men and women. To account for multiple measures from the same person (both knees, multiple visits), Generalized Estimating Equations (GEE) were used to investigate the relationship between shape, KLG and changes between visits after adjustment for age, sex and BMI (SPSS v22, IBM corp). Both knees were included at each time-point and the dependent variable was the mode of interest. A separate GEE was run for each mode. Variables analysed were Baseline KLG, Visit (in months), age at baseline, sex, BMI and the interaction between visit and baseline KLG. An autoregressive correlation matrix (AR(1)) was used. The distribution was set to 'normal' and link to 'identity'. GEE generates a measure of the slope of the relationship, B, and its 95% confidence interval. The associated *P*-value was used merely as a guide and we took results to be statistically significant when the 95%CI did not contain zero.

Where statistically significant changes over time (*i.e.* with visit) were observed, SRMs were calculated using the knee with the highest KLG for each person (MedCalc Software, version 15, Ostend, Belgium). The RC index was calculated for each idividual over each period by subracting their mode score at the earlier timepoint from that at the subsequent timepoint and dividing the difference by the standard error of the measurements given by

$$SE = \sqrt{S_1^2 + S_2^2 - 2S_1S_2r_{12}}$$

where S_1 and S_2 are the standard deviations of the scores at the first and second time points and r_{12} is the Pearson's correlation coefficient between the two sets of scores [11]. For each individual, the RC calculation generates a *z*-score and we follow the convention that a *z*-score greater in magnitude than 1.96 denotes a statistical significance at *P*<0.05 to identify a result that is likely real rather than a consequence of random variation in the measurements. For each individual between each pair of time points the RC index identifies changes in mode score that can be expressed in one of three categories: increased (RC ≥ 1.96), decreased (RC ≤ -1.96) or unchanged (-1.96 < RC < 1.96). A threshold for a change in mode score between two time points that may be considered statistically reliable may be calculated from 1.96*SE* [11].

RESULTS

In total, 109 subjects (60 female and 49 male) were recruited and had baseline images assessed. Of these, 101 had measurements at 6 months and 78 at 12 months. At baseline, 57 knees were classed as KLG0, 80 as KLG1, 47 as KLG2, 26 as KLG3 and 8 as KLG4 from

the radiograph taken in the preceding 12 months. The mean time between baseline and the 6-month DXA scan was 169 (22) days, between 6- and 12-month DXA scans was 182 (13) days and for baseline to 12-month DXA scan was 350 (30) days. Table 1 shows the numbers in each group and the results of tests for differences in the distributions of age, sex and BMI between the four severity groups.

The first five modes of variation were selected for analysis from the scree plot and together these modes explained 43% of the total variance. Mode 5 described 4.7% of the variance in the model and subsequent excluded modes were all smaller. Using GEE, statistically significant associations with KLG, after adjustment for age, sex and BMI, were observed in two modes; mode 1 (B=0.38 [0.28, 0.49], P<0.0001) and mode 3 (B=–0.30 [–0.47, –0.12], P<0.001). Whilst mode 1 did not change significantly during the study period, mode 3 showed significant differences with visit (B=–0.23 [–0.32, –0.14], P<0.0001) (Table 2). Further investigation including an interaction term between KLG and visit showed a small but significant interaction for mode 3 (B=–0.094 [–0.16, –0.025], P=0.007), but no interaction for mode 1 (P=0.16). Both modes associated with KLG were significantly associated with age and mode 1 was also related to BMI (P<0.01). Two modes were associated with sex (Modes 2 and 4 (P<0.05)) but were not associated with any other input variables in the model.

Figure 2 shows the variations in shape described by modes 1 and 3 and the association with KLG is shown in Supplementary Figure S2. Increasing Mode 1 values show the symmetrical development of classical features of radiographic OA with femoral and tibial osteophytes growing as the joint space between the femur and tibia reduces and the tibial plateau widens. Mode 3 scores were negatively associated with increasing radiographic severity and, like mode 1, described increasingly large osteophytes and joint space narrowing. Unlike mode 1, this was asymmetrical with narrowing and tibial osteophyte formation occurring primarily on the medial side with decreasing scores. The largest change, however, was in alignment of the femur and tibia; decreasing mode 3 scores indicated increasing levels of displacement with the intercondylar notch offset from the tibial spines and the lateral tibial plateau projecting beyond the femur.

Progressive changes in mode 3 scores were observed overall and especially in the moderate and severe groups, with larger changes occurring in the first 6 months than in the second. SRM values were calculated from the knee with the highest KLG at baseline (used for severity grouping) for the whole 12 months, and for each successive 6 month period (Table 3). Overall, the 12-month SRM for mode 3 was 0.63. When this was analysed by severity group the largest change was seen in those with moderate OA (12-month SRM 0.92), followed by those with

doubtful or severe OA (12-month SRMs 0.69 and 0.70 respectively). For completeness, the mean of the changes in mode 3 over each period are shown in Supplementary Table S1.

For individual participants, considering the more severely affected knee, between baseline and 6 months 9 of the 101 participants (8.9%) had an RC index that exceeded -1.96 and 1 had an RC index greater than 1.96. Between baseline and 12 months 10 of the 78 remaining participants (12.8%) had an RC index exceeding -1.96 and one had an RC index greater than 1.96. Between 6 months and 12 months only one had an RC index greater than -1.96 whereas 3 had an RC index greater than 1.96. When analysed by severity group the most reliable changes were seen in the moderate and doubtful groups with 14% of participants (6/43) having an RC index exceeding -1.96 and none whose RC index was above 1.96 over the 12 months from baseline. Over a 12-month period the magnitude of the reliable change threshold was calculated to be a change in mode 3 score of 1.23 in the whole cohort and 1.24 in the doubtful/moderate groups combined.

DISCUSSION

This study showed that statistical shape modelling from DXA images can be used to model the shape changes in the knee that occur with OA and that it is sensitive enough to identify morphological changes over a 12-month period. Overall, for the whole cohort, the SRM value of -0.63 for mode 3 compares favourably with SRM values reported for radiographic joint space width (the only measure of OA progression currently recommended for clinical trials by the FDA), which are typically in the region of 0.13 to 0.37 (95%CI) for radiographic measures over 1-2 years [2]. When examined by severity group, SRM values ranged from -0.7 for mild and severe OA and greater than -0.9 for moderate OA (KLG2) over 12 months. Other measures have reported higher SRM values but have not been approved for clinical trials, for example the use of ordered values of cartilage thickness [4], measurement of contact area between the femur and tibia (SRM up to 1.14 in those with severe OA over 21 months) [30], or bone area (SRM 0.83 over 2 years) [31].

The reliable change index is an attempt to identify and quantify observations that are clinically significant [8]. It was recently adopted to distinguish true changes in joint space width (JSW) from measurement error in individuals with knee OA [11]. In their study of 559 individuals, 70% had an apparent reduction, and 30% an increase, in JSW over the first 12-month period. Using the RC index, however, showed that only 6% had a statistically reliable reduction in JSW and 1.1% showed an increase [11]. Over a 12 month period using SSM we found 12.8% of the whole cohort had a statistically reliable reduction in mode 3 score and, if we limited the

analysis to those with KLG of 1 and 2, our doubtful and moderate groups, this increased slightly to 14%.

The changes in shapes described by these modes are consistent with traditional descriptions of OA development. Mode 1 shows the "classical" signs of OA displaying symmetrically, which did not change significantly during the course of the study, whereas mode 3 showed asymmetrical changes. Angular malalignment, particularly a varus (bow-legged) deformity, is strongly linked to OA and whilst several studies find associations with OA severity or progression [32-34] its role as a risk factor for incident OA is not yet clear [35, 36]. Although we cannot measure angular malalignment, the changes we observed in mode 3 demonstrate an increasing offset between the femur and tibia along with with medial osteophytes and joint space narrowing.

Mode 3 variation was evident in those with existing radiographic OA, and largest in those with moderate OA at baseline (KLG2). Those with severe radiographic OA are a particularly difficult group in whom to monitor progression as measures such as JSW become vanishingly small. It can be seen in Supplementary Information that mode scores often do not vary linearly with KLG and, for mode 3 from baseline radiographs, it can be seen to change little unitl OA is severe (KLG 3 and 4). This possibly explains why mode 3 scores at baseline appear to increase with severity rather than decrease, although the 95% confidence interval encompasses the scores for control to moderate OA. Here, though, we show that SSM can measure changes in people with severe OA comparable with those found in subjects with mild OA. Although some aspects of these shapes, *e.g.* osteophytosis, could be detected by modelling a single bone, by looking at the femur and tibia together we can measure changes in the whole joint. Others, such as Neogi et al., who used 3D SSM to predict TKR, also found that incorporating multiple bones can improve the performance of an SSM [19], although 3D models are considerably more complex.

The SRM for mode 3 differed between the two time intervals. It changed rapidly in the first 6 months then only a little during the following 6 months. Whilst this was unexpected it is not unusual [4, 37]. To ensure these results were not influenced by environmental factors, we re-ran the analysis, adjusting for systematic effects that could have influenced the results, including the number of days between scans (not just months to each visit), which radiographer performed each scan and the temperature in Aberdeen on their scan date but none of these had either a significant or noticeable effect on the results. Positioning of the individuals could have played a part, but this was done by the same radiographers on the same instrument and unlikely to be consistenly different at each visit. Use of a model based

9

on a larger reference cohort may help to eliminate any bias that may have arisen from systematic factors we have been unable to account for and work is in hand to establish such a reference model. With no identifiable systematic source of this variation, use of 12-month data seems more reliable for assessment of the knee SSM in this study but a shorter timescale could be addressed in future studies.

The strengths of this study include the short time between repeated scans and the wide range of osteoarthritic severity in the cohort. Testing of this model using DXA scans was of particular benefit as the scans are quick and the scanner is of open design. Whilst DXA scanners vary greatly in resolution, images from the iDXA (GE Lunar) are of near-radiographic quality and a recent study has shown that radiographic OA can be graded as accurately using iDXA as on standard radiographs [38]. This makes DXA an attractive imaging modality for investigation of musculoskeletal disorders. For use as a biomarker in large trials these advantages are not inconsiderable and in addition, the technology is relatively inexpensive. While MRI has several advantages, such as being able to generate a 3D model that removes positioning problems and identifying localised changes in joint shape and internal features, as a biomarker it has a number of limitations associated with fitting large or claustrophobic people inside an enclosed scanner or requiring them to remain still for prolonged periods, despite stiff and painful joints. The associated limitation of DXA, however, is that these models were restricted to 2-dimensions. This may result in some positioning errors, especially as OA gets more severe and internal rotation at the hip becomes more pronounced. However, positioning is done in a standardised way by trained radiographers and this might mitigate some of the possible errors due to different positioning with each visit. Whilst the recruitment process, using database searches of existing radiographs taken during routine clinical practice was efficient and cost effective, it also introduced some limitations in that selection and stratification of the study was based on radiographic measures alone and that the time between that radiograph and the baseline DXA after recruitment to the study varied by several months [38]. Numbers in our proof-of-concept study are small but these data indicate that a larger study is called for. If these figures are confirmed then RC index measures would indicate that a only half the number of participants might be required in a study to match results based on JSW. The background and limitations of the RC technique related to use in musculoskeletal studies are discussed in more detail by Parsons et al. [11]. The different RC variations that have been proposed have been compared in detail by Hinton-Bayre and there is presently no consensus as to which RC index methodology should be used [10]. A criticism of the RC index is that it is specific but not very sensitive due, in part, to the magnitude of the measurement errors associated with longitudinal studies. We used the common convention of a 5% level of significance, meaning that the cut-point for RC index scores was ±1.96. This is clearly an

arbitrary threshold and the sensitivity of the RC index could be increased by using a less strict cut-off [11].

In conclusion, this novel study shows that statistical shape modelling of the knee can be a sensitive marker of structural progression in osteoarthritis in longitudinal studies over a period of 12 months. In addition, although similar results may be expected from carefully positioned radiographs, we have shown that this can be performed using DXA scans, thereby incurring a much lower radiation dose.

References

1   Duryea J, Neumann G, Niu J, et al. Comparison of radiographic joint space width with magnetic resonance imaging cartilage morphometry: analysis of longitudinal data from the Osteoarthritis Initiative. Arthritis Care Res.(Hoboken.) 2010;62(7):932-7.

2   Reichmann WM, Maillefert JF, Hunter DJ, Katz JN, Conaghan PG, Losina E. Responsiveness to change and reliability of measurement of radiographic joint space width in osteoarthritis of the knee: a systematic review. Osteoarthritis Cartilage 2011;19(5):550-6.

3   Kinds MB, Marijnissen ACA, Bierma-Zeinstra SMA, et al. Detection of progression of radiographic joint damage in case of very early osteoarthritis: Sensitivity to change of quantitative analysis compared to qualitative grading. Rheumatology Reports 2012;4(1):27-33.

4   Wirth W, Larroque S, Davies RY, et al. Comparison of 1-year vs 2-year change in regional cartilage thickness in osteoarthritis results from 346 participants from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2011;19(1):74-83.

5   Eckstein F, Guermazi A, Gold G, et al. Imaging of cartilage and bone: promises and pitfalls in clinical trials of osteoarthritis. Osteoarthritis Cartilage 2014;22(10):1516-32.

6   Hunter DJ, Zhang W, Conaghan PG, et al. Responsiveness and reliability of MRI in knee osteoarthritis: a meta-analysis of published evidence. Osteoarthritis Cartilage 2011;19(5):589-605.

7   Norman GR, Wyrwich KW, Patrick DL. The mathematical relationship among different forms of responsiveness coefficients. Qual Life Res 2007;16(5):815-22.

8   Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J. Consult. Clin. Psychol. 1991;59(1):12-9.

9   Hinton-Bayre AD. Deriving reliable change statistics from test-retest normative data: comparison of models and mathematical expressions. Arch. Clin. Neuropsychol. 2010;25(3):244-56.

10   Hinton-Bayre AD. Clarifying Discrepancies in Responsiveness Between Reliable Change Indices. Arch. Clin. Neuropsychol. 2016;31:754-68.

11   Parsons C, Judge A, Leyland K, et al. Novel approach to estimate Osteoarthritis progression - use of the reliable change index in the evaluation of joint space loss. Arthritis Care Res. (Hoboken) 2018;0(ja).

12   Gregory JS, Waarsing JH, Day J, et al. Early identification of radiographic osteoarthritis of the hip using an active shape model to quantify changes in bone morphometric features - Can hip shape tell us anything about the progression of osteoarthritis? Arthritis and Rheumatism 2007;56:3634-43.

13   Faber BG, Baird D, Gregson CL, et al. DXA-derived hip shape is related to osteoarthritis: findings from in the MrOS cohort. Osteoarthritis Cartilage 2017;25(12):2031-8.

14   Barr RJ, Gregory JS, Yoshida K, Alesci S, Aspden RM, Reid DM. Significant morphological change in osteoarthritic hips identified over 6-12 months using statistical shape modelling. Osteoarthritis Cartilage 2018;26(6):783-9.

15   Barr RJ, Gregory JS, Reid DM, et al. Predicting osteoarthritis progression to total hip replacement: can we do better than risk factors alone using active shape modelling as an imaging biomarker? Rheumatology (Oxford) 2012;51(3):562-70.

16   Agricola R, Reijman M, Bierma-Zeinstra SMA, Verhaar JAN, Weinans H, Waarsing JH. Total hip replacement but not clinical osteoarthritis can be predicted by the shape of the hip: A prospective cohort study (CHECK). Osteoarth Cartil 2013;21(4):559-64.

17   Nelson AE, Liu F, Lynch JA, et al. Association of incident symptomatic hip osteoarthritis with differences in hip shape by active shape modeling: The johnston county osteoarthritis project. Arthritis Care and Research 2014;66(1):74-81.

18   Lynch JA, Parimi N, Chaganti RK, Nevitt MC, Lane NE. The association of proximal femoral shape and incident radiographic hip OA in elderly women. Osteoarth Cartil 2009;17(10):1313-8.

19   Neogi T, Bowes MA, Niu J, et al. Magnetic Resonance Imaging-Based Three-Dimensional Bone Shape of the Knee Predicts Onset of Knee Osteoarthritis: Data From the Osteoarthritis Initiative. Arthritis Rheum. 2013;65(8):2048-58.

20   Bredbenner TL, Eliason TD, Potter RS, Mason RL, Havill LM, Nicolella DP. Statistical shape modeling describes variation in tibia and femur surface geometry between Control and Incidence groups from the Osteoarthritis Initiative database. Journal Of Biomechanics 2010;43(9):1780-6.

21   Haverkamp DJ, Schiphof D, Bierma-Zeinstra SM, Weinans H, Waarsing JH. Variation in joint shape of osteoarthritic knees. Arthritis Rheum. 2011;63(11):3401-7.

22   Varzi D, Coupaud SA, Purcell M, Allan DB, Gregory JS, Barr RJ. Bone morphology of the femur and tibia captured by statistical shape modelling predicts rapid bone loss in acute spinal cord injury patients. Bone 2015;81:495-501.

23   The Atlas of Standard Radiographs of Arthritis. Rheumatology 2005;44(suppl 4):iv43-iv72.

24   Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann. Rheum. Dis. 1957;16:494-502.

25   Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159-74.

26   Cootes TF, Hill A, Taylor CJ, Haslam J. Use of active shape models for locating structures in medical images. Image and Vision Computing 1994;12:355-65.

27   Cattell RB. The Scree Test For The Number Of Factors. Multivariate Behav Res 1966;1(2):245-76.

28   Cohen J. *Statistical power analysis for the behavioral sciences*: Lawrence Erlbaum Associates; 1988.

29   Mazzuca SA, Brandt KD, Buckwalter KA, Lequesne M. Pitfalls in the accurate measurement of joint space narrowing in semiflexed, anteroposterior radiographic imaging of the knee. Arthritis Rheum. 2004;50(8):2508-15.

30   Tummala S, Nielsen M, Lillholm M, Christiansen C, Dam EB. Automatic quantification of tibio-femoral contact area and congruity. IEEE Trans. Med. Imaging 2012;31(7):1404-12.

31   Bowes MA, Vincent GR, Wolstenholme CB, Conaghan PG. A novel method for bone area measurement provides new insights into osteoarthritis and its progression. Ann Rheum Dis 2015;74(3):519-25.

32   Kinds MB, Marijnissen AC, Vincken KL, et al. Evaluation of separate quantitative radiographic features adds to the prediction of incident radiographic osteoarthritis in individuals with recent onset of knee pain: 5-year follow-up in the CHECK cohort. Osteoarthritis Cartilage 2012;20(6):548-56.

33   Kinds MB, Marijnissen AC, Viergever MA, Emans PJ, Lafeber FP, Welsing PM. Identifying phenotypes of knee osteoarthritis by separate quantitative radiographic features may improve patient selection for more targeted treatment. J Rheumatol 2013;40(6):891-902.

34   Brouwer GM, van Tol AW, Bergink AP, et al. Association between valgus and varus alignment and the development and progression of radiographic osteoarthritis of the knee. Arthritis Rheum 2007;56(4):1204-11.

35   Hunter DJ, Niu J, Felson DT, et al. Knee alignment does not predict incident osteoarthritis: the Framingham osteoarthritis study. Arthritis Rheum 2007;56(4):1212-8.

36   Zhai G, Ding C, Cicuttini F, Jones G. A longitudinal study of the association between knee alignment and change in cartilage volume and chondral defects in a largely non-osteoarthritic population. J Rheumatol 2007;34(1):181-6.

37   Raynauld JP, Martel-Pelletier J, Berthiaume MJ, et al. Long term evaluation of disease progression through the quantitative magnetic resonance imaging of symptomatic knee

osteoarthritis patients: correlation with clinical symptoms and radiographic changes. Arthritis Res Ther 2006;8(1):R21.

38   Yoshida K, Barr RJ, Galea-Soler S, Aspden RM, Reid DM, Gregory JS. Reproducibility and Diagnostic Accuracy of Kellgren-Lawrence Grading for Osteoarthritis Using Radiographs and Dual-Energy X-ray Absorptiometry Images. Journal of Clinical Densitometry 2015;18(2):239-44.

Competing interests

Dr Salvatore Alesci was an employee of Wyeth Inc., one of the funders of TMRC, at the time of the study. All other authors have no competing interests to declare.

Table 1. Distribution of sex, BMI and age for each severity group at baseline. Significance differences between groups were tested using Chi Squared (sex), one-way ANOVA (BMI). Age was not normally distributed (Shapiro-Wilk test = 0.025), so Kruskal-Wallis ANOVA on Ranks was used. Significant differences were found between the severity groups for age, sex and BMI.

Severity Group	N	Sex N female (%)	BMI Mean (st. dev.)	Age Median (25% - 75%)
Control	37	21 (56.8%)	27.7 (4.2)	44.8 (40.7, 56.1)
Doubtful	24	13 (54.2%)	28.8 (4.8)	64.5 (56.5, 70.3)
Moderate	24	18 (75%)	31.1 (4.0)	67.8 (56.6, 73.8)
Severe	24	8 (33.3%)	32.4 (5.8)	67.0 (61.9, 75.1)
Total	109	60 (55%)	29.8 (5.0)	61.5 (52.6, 69.9)
P-value		0.037	< 0.001	< 0.001

Table 2. Scores and 95%CI for mode 3 in each severity group at each visit

| Group	Baseline	Month 6	Month 12
Control	0.18 [-0.11, 0.48]	-0.07 [-0.36, 0.22]	-0.35 [-0.79,0.09]
Doubtful	0.33 [-0.04, 0.69]	-0.03 [-0.42, 0.37]	-0.05 [-0.42, 0.31]
Moderate	0.54 [0.15, 0.94]	-0.04 [-0.34, 0.43]	-0.04 [-0.43, 0.35]
Severe	-0.21 [-0.82, 0.40]	-0.77 [-1.35, -0.18]	-1.05 [-1.79, -0.32]
All groups	0.21 [0.01, 0.41]	-0.17 [-0.37, 0.02]	-0.34 [-0.58, -0.10]

Table 3. Resposiveness of mode 3 for each severity group. Standardized Response Means (SRM) and 95% confidence intervals of changes between timepoints based on the knee with the highest KLG.

| Group	Months 0-12	Months 0-6	Months 6-12
Control	-0.21 [-0.71, 0.28]	-0.42 [-0.77, -0.04]	0.10 [-0.44, 0.58]
Doubtful	-0.69 [-1.07, -0.29]	-0.64 [-1.1, -0.13]	-0.05 [-0.49, 0.43]
Moderate	-0.92 [-1.28,-0.58]	-0.91 [-1.30, -0.49]	-0.25 [-0.68, 0.25]
Severe	-0.70 [-1.25, -0.1]	-0.56 [-1.20, -0.02]	-0.14 [-0.58, 0.43]
All groups	-0.63 [-0.87, -0.40]	-0.59 [-0.79, -0.37]	-0.08 [-0.29, 0.16]

Figure captions

Figure 1. Typical DXA image and corresponding template for SSM of the knee. (a) Image of a left knee obtained from iDXA (GE Lunar) scanner showing the points marking the outline from the statistical shape model. Key points (marking definable anatomical features) are in red, the remainder in yellow, although all points are treated the same way in analysis. (b)Template for the knee SSM comprising the distal femur, proximal tibia and osteophytes. Colours are used to highlight each part of the template. The femur is shown in blue and the tibia in green with the rear of the tibial plateau in olive. The femoral medial and lateral osteophytes are shown in purple (left) and cyan (right) and those of the tibia in orange and red. In the absence of osteophytes these lines collapse onto the main outline.
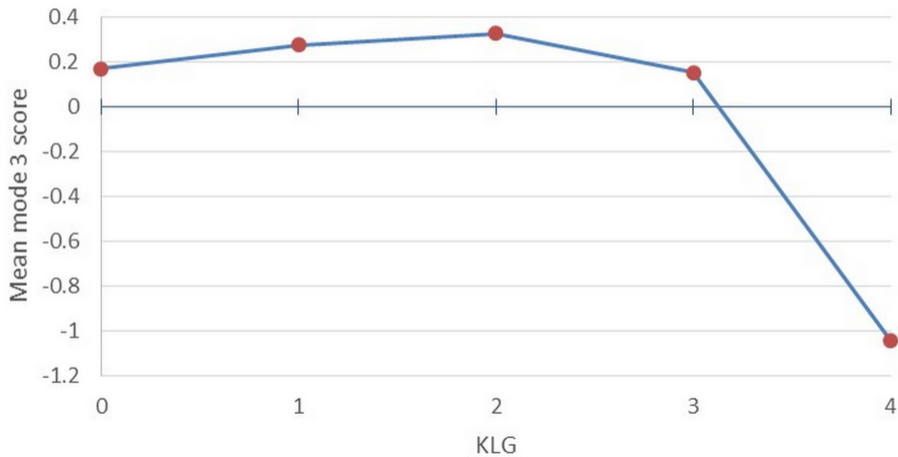
Figure 2. Shapes described by ±2 standard deviations of modes 1 and 3. Mode 1 identifies symmetrical development of osteophytes and loss of joint space with positive scores (solid red line) and Mode 3 describes medial osteophyte development and increasing displacement with lower scores (dashed blue line).
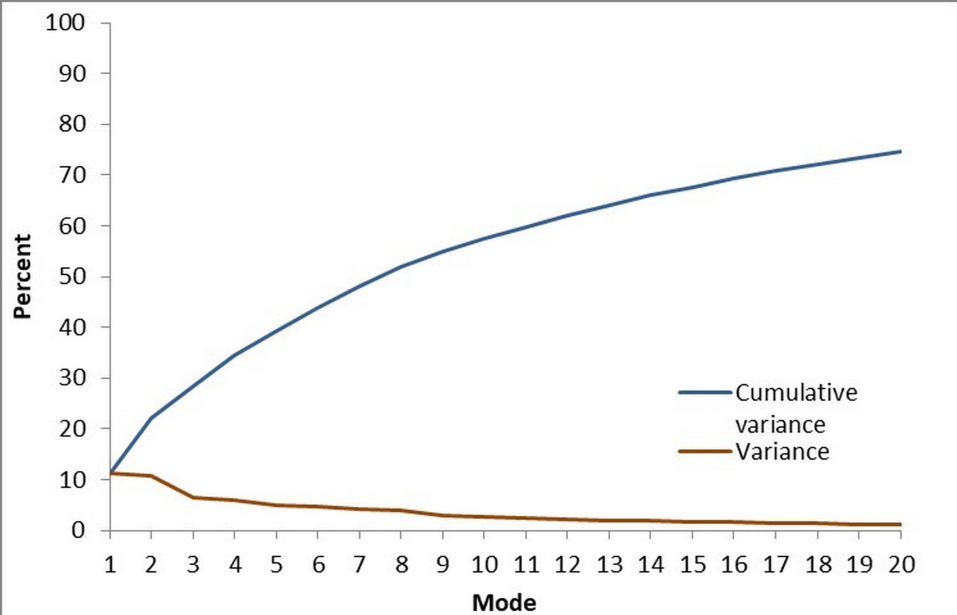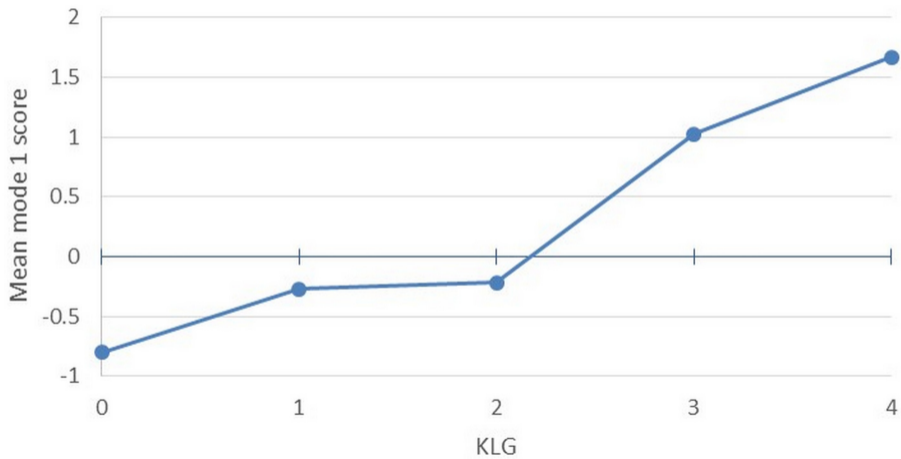
Mode 3

Statistical Shape Modelling provides a responsive measure of morphological change in knee osteoarthritis over 12 months.

Jennifer S Gregory, Rebecca J Barr, Kanako Yoshida, Salvatore Alesci, David M Reid, Richard M Aspden

Supplementary Information

Supplementary Data S1.

Detailed Methods

Imaging

Knee radiographs were taken using postero-anterior, weight-bearing, semi-flexed metatarsophalangeal (MTP) views described by Buckland-Wright et al. (Buckland-Wright et al., 2004, Buckland-Wright et al., 1999). A single cassette size of 30 x 40 cm was used, and the source-film distance was 120 cm with exposure factor of 55 kV, 4 mA s for a standard subject.

Dual Energy X-ray Absorptiometry (DXA) images were acquired according to local standard protocol for 'dual proximal femur' using imaging DXA (Lunar, GE) by experienced senior radiographers specialising in DXA. Each knee was scanned separately, with the knee fully extended and positioned as in standard non-weight-bearing knee radiographs. A scout scan targeting the femorotibial joint space was performed initially in the majority of patients, for optimal positioning of the knee to visualise the tibial spines. Scanning was initiated 4 cm above the apex of the patella.

Statistical Shape Modelling

Statistical Shape Modelling (SSM) is a statistical image analysis method for quantifying the variation in a set of similar shapes. Details of the statistical shape modelling technique (SSM) used have been described elsewhere (Cootes et al., 1994). The method was first developed for use in image segmentation; a method for identifying objects in a complex image. An SSM template is a set of landmark points that define the shape to be identified. To aid comparison between shapes each point is always placed on the same feature of the outline. The Nottingham line drawing atlas (Nagaosa et al., 2000) was used as the basis for the template for the SSM of the knee. After training no a small subset of images, the search is automatic to identify the outline. However, all the fitted points are checked manually and corrected, if necessary, to ensure the outlines of each part have been correctly identified. In all our studies of repeated point placement, final point placement, both intra-and inter-observer, is typically within a few pixels.

The method works by calculating in each image the distance of the set of landmark points marking the outline of the object of interest from the mean position of equivalent points calculated from a set of images. This is done following a set of affine transformations (Procrustes analysis) to ensure that all objects in the dataset have been aligned as closely as possible, so differences in point position are due to genuine differences in shape rather than size, location or translation. Key points are placed at easily identifiable features whilst the remaining points are spaced approximately evenly between them. The shape is then described by a series of orthogonal 'modes of variation' which are derived using a statistical method called Principal Components Analysis. This approach enables the dataset to be described in terms of a smaller number of variables called principal components. This is somewhat analogous to Fourier analysis of a tone that enables a complicated sound to be described in terms of harmonics, or a series of notes each with a single frequency. For the technically minded, these modes, or principal components, are the eigenvalues and eigenvectors of the covariance matrix derived from the point coordinates as described above.

The modes of variation are ranked in decreasing order of the amount of variance described; the first few modes generally account for the majority of the variance within the dataset, whilst the later modes account for little and can usually be neglected as

noise. The amount of variance described by each mode may be plotted as a scree plot (Figure S1). Each mode of variation is normalized to have zero mean and unit standard deviation and each image can then be described in terms of how many standard deviations each mode of variation lies from the average. Each mode is orthogonal to, *i.e.* linearly independent of, all the others (although there may still be non-linear relationships).

Because of the normalization, overall size and orientation are removed as scaling factors enabling the actual femoral morphologies to be compared. This overcomes many of the problems caused by the strong correlations found between different geometrical measurements.

Knee model

In this study we built several SSMs, to include the distal femur, proximal tibia and osteophytes, to explore various point placements This enabled visualization of common radiographic features observed in OA. An early template was used as a pilot study and applied to the baseline radiographs to show that mode scores do correspond with radiographic severity assessed using KLG. This model, however, included a point at the apex of the patella which was found to be difficult to position reproducibly, so this point was removed and the template modified slightly before being used on the DXA images at the three time-points. It is difficult to compare change in mode score with change in KLG as mode scores are quantifiable, continuous variables whereas KLG are subjective, visual gradings based on radiographic features and do not have the same sensitivity, *e.g.* a grade of 2.7 is not possible. This indicates how SSM has the potential to be much more sensitive to subtle changes in joint shape reflecting OA than a grading scheme such as KLG. Figure S1 shows a scree plot of the variance described by each of the first 20 modes against mode number. Modes 1-5 of the DXA model were chosen for further analysis
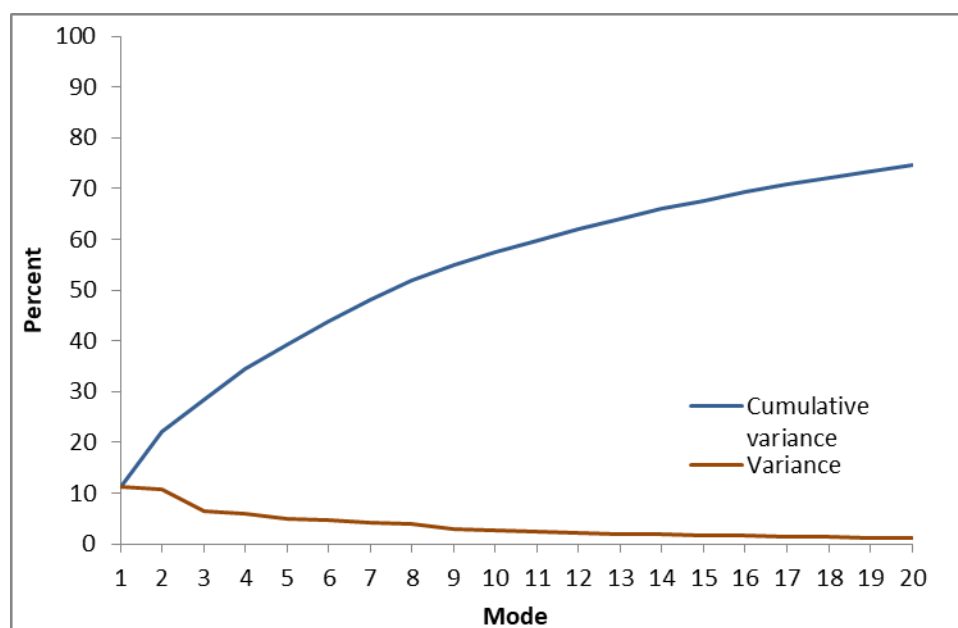
Figure S1. **A scree plot of variance explained by each mode**. Plots show the variance and cumulative variance for the first 20 modes. Choice of how many to analyse further is guided by looking for points at which the gradient changes.

References

Buckland-Wright, JC, Ward, RJ, Peterfy, C, Mojcik, CF, Leff, RL (2004) Reproducibility of the semiflexed (metatarsophalangeal) radiographic knee position and automated measurements of medial tibiofemoral joint space width in a multicenter clinical trial of knee osteoarthritis. *The Journal of Rheumatology* **31,** 1588-1597.

Buckland-Wright, JC, Wolfe, F, Ward, RJ, Flowers, N, Hayne, C (1999) Substantial superiority of semiflexed (MTP) views in knee osteoarthritis: a comparative radiographic study, without fluoroscopy, of standing extended, semiflexed (MTP), and schuss views. *J Rheumatol* **26,** 2664-2674.

Cootes, TF, Hill, A, Taylor, CJ, Haslam, J (1994) Use of active shape models for locating structures in medical images. *Image and Vision Computing* **12,** 355-365.

Nagaosa, Y, Mateus, M, Hassan, B, Lanyon, P, Doherty, M (2000) Development of a logically devised line drawing atlas for grading of knee osteoarthritis. *Ann Rheum Dis* **59,** 587-595.

Supplementary Data S2.

Knee Shape and Kellgren-Lawrence grading

The SSM for the knee was applied to all the DXA images from both knees of all individuals enrolled in the study at all three time points. The average knee shape is then calculated. All of the images are included in the SSM because it has to describe the whole of the variation in the shapes present in the dataset. As a consequence, the mean shape from the SSM is effectively an outline of a knee with moderate OA. Correlations between each knee-shape mode at the baseline visit and KL grade were assessed to identify modes likely to be of interest for assessing OA progression. (Figure S2).
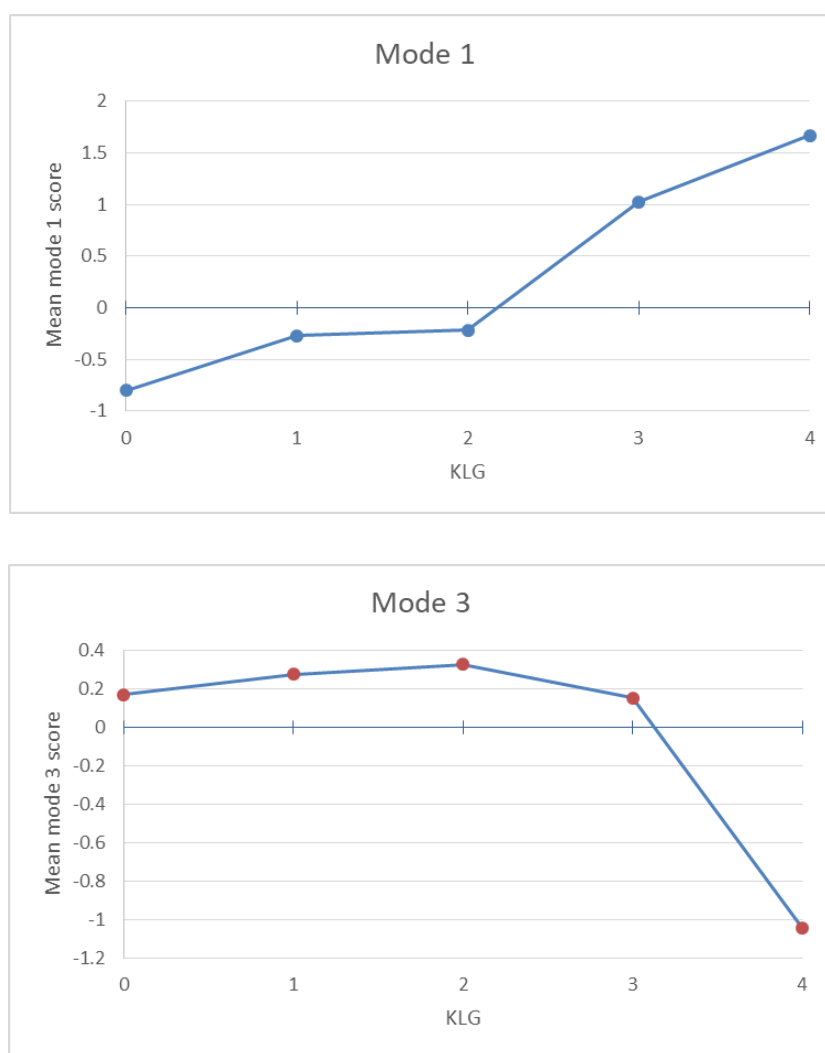
Figure S2. Plots of the mean score for Shape modes 1 and 3 at each KL grade. An increase in KL grade is associated with an increased mode 1 score in which evidence of femoral and tibial osteophyte formation and uniform joint space narrowing in both compartments is seen. Mode 3 shows little change at low KL grades but decreases rapidly at higher KLG. Medial osteophyte formation and joint space narrowing are associated with low (negative) mode scores.

Supplementary Data S3.

The standardized response mean (SRM) is calculated from the mean change in scores divided by the standard deviation of the change. Note this is not the same as the change in the mean scores shown in Table 2. For completeness, we include here the data used in the calculations of SRM values for mode 3 shown in Table 3 of the main article.

Table S1. Mode 3 scores for each severity grade. Changes shown as the mean of the change in mode 3 scores, the SD of the change and the number of participants, N, remaining at the end of each period.

Group		Months 0-12	Months 0-6	Months 6-12
Control	Change	-0.13	-0.27	0.047
	SD	0.63	0.63	0.47
	N	18	34	18
Doubtful	Change	-0.43	-0.35	-0.033
	SD	0.62	0.55	0.63
	N	23	24	23
Moderate	Change	-0.57	-0.53	-0.10
	SD	0.62	0.58	0.38
	N	20	23	20
Severe	Change	-0.46	-0.53	-0.10
	SD	0.65	0.95	0.72
	N	17	20	17
All groups	Change	-0.40	-0.40	-0.045
	SD	0.64	0.68	0.55
	N	78	101	78